Probability in the Engineering and Informational Sciences, **27**, 2013, 309–317.
doi:10.1017/S0269964813000041

A TIME-DEPENDENT STUDY OF THE KNOCKOUT QUEUE

BRIAN FRALIX
Department of Mathematical Sciences
Clemson University
Clemson, SC
E-mail: bfralix@clemson.edu

We examine the time-dependent behavior of a birth–death process, whose birth rates and death rates are decreasing and increasing, respectively, with respect to the current state. Such models can be used to describe Markovian queueing systems with exponential reneging, where potential arrivals balk with a certain probability that depends on the number of customers observed upon arrival. Our results are derived by interpreting the birth–death process as the queue-length process of what we refer to as the "knockout queue."

1. INTRODUCTION

In most real-life queueing systems, a customer has the opportunity to balk (i.e., decide not to enter the system) upon observing the state of the system at the time of arrival. Moreover, a customer that does enter the system may choose to renege if he is unsatisfied with the quality of service received while in the queue. It stands to reason then that when forming a mathematical model of a queueing system, allowing customers within the model to balk or renege can tremendously improve our ability to accurately describe the queueing system.

We present a time-dependent analysis of a birth–death process $\{Q(t); t \geq 0\}$ on the state space $\{0, 1, 2, \ldots, N\}$, with N being either finite or infinite. The birth rates are decreasing up to state N, that is $\lambda_n \geq \lambda_{n+1}$ for $0 \leq n \leq N-1$, with $\lambda_n = 0$ for $n \geq N$, and the death rates are increasing in n, so that $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_N$. The sojourn time in each state n is exponential with rate $\lambda_n + \mu_n$ for $n \geq 1$, and λ_0 for $n = 0$.

This framework allows us to consider Markovian queueing models that allow for customer balking or reneging to occur. An example of such a model is a system where arrivals that observe n customers in the system decide to join upon arrival with probability p_n, and where customers that do join the system bring unit exponential amounts of work to a single server operating at rate μ that processes work in a first-come-first-served (FCFS) manner, but are only willing to wait in the system for at most an exponentially distributed patience time with parameter γ. The birth and death parameters corresponding to such a system are as follows: $\lambda_n = \lambda p_n$, and $\mu_n = \mu + (n-1)\gamma$, for $0 \leq n \leq N$ (with $\lambda_N = \mu_0 = 0$).

In fact, it is always possible to associate a simple queueing interpretation to $\{Q(t); t \geq 0\}$. Indeed, this process represents the queue-length process of a FCFS queueing system with

© Cambridge University Press 2013 0269-9648/13 $25.00

https://doi.org/10.1017/S0269964813000041 Published online by Cambridge University Press

balking, where potential customers arrive to the system in accordance to a homogeneous Poisson process with rate λ_0, and enter the system with probability λ_n/λ_0 whenever an arrival observes n customers present in the system upon arrival (otherwise they balk from the system). Moreover, each arrival entering the system brings a unit exponential amount of work for processing, where the server processes work at rate μ_n whenever there are n customers present in the system. We refer to this system throughout the paper as the FCFS system.

By constructing another queueing system whose queue-length process is equal in distribution to Q, we are able to compute expressions for both the first moment of $Q(t)$, $t \geq 0$, as well as the time-dependent balking probability of the FCFS queueing system. We refer to our constructed queueing system as the *knockout queue*, or *knockout model*. The knockout queue has many appealing properties that make it suitable for a time-dependent study via the transient Little's law, which is discussed in [8]. The techniques used here are very similar to those found in both [7] and [8], yet the results given here also depend heavily on the deliberate construction of our knockout model, which seems not to have been studied until now. The expressions we derive are in terms of events associated with hitting-times of a continuous-time Markov chain (CTMC) with absorbing states, and are analogous in form to the results of Abate and Whitt [1–4] for the $M/M/1$ queue and for regulated Brownian motion, along with the results found in [7,8].

We also illustrate how our results simplify for two classical models: the $M/M/N/N$ loss model, and the $M/M/1/N$ model. The transient behavior of the loss model is an object of interest in many studies: see Abate and Whitt [5] and (with a different approach) Virtamo and Aalto [15] (with both studies focusing on exact calculations of the first moments and/or loss probabilities), along with Mandjes and Ridder [12], who focus on deriving various types of simple approximations for the loss probabilities.

2. CONSTRUCTING THE KNOCKOUT QUEUE

We now construct the knockout queue. We assume $N < \infty$ throughout, as the case where $N = \infty$ is simpler, and less interesting.

Suppose customers arrive to the system in accordance to a Poisson process with rate λ_0, and each customer brings with it an exponential amount of work with unit rate, independently of all other customers. This queueing system consists of N slots, where customers always occupy the lowest available slots: for example, when there are n customers in the system, they can be found among the first n slots.

When a customer arrives to the system, he first decides whether or not to eliminate, or "knockout" one (and only one) of the customers currently found in the queue. Assuming there are n customers present in the system upon arrival, the arriving customer eliminates the customer found in slot k with probability α_k, $1 \leq k \leq n$, where

$$\alpha_k = \left(1 - \frac{\lambda_k}{\lambda_{k-1}}\right) \prod_{i=1}^{k-1} \frac{\lambda_i}{\lambda_{i-1}} = \frac{\lambda_{k-1} - \lambda_k}{\lambda_0},$$

or decides not to eliminate any customers with probability λ_n/λ_0. If the arrival eliminates the customer in slot k, he joins slot 1, and all customers formerly in slot i move to slot $i+1$, $1 \leq i \leq k-1$. Otherwise, if the arrival does not eliminate anyone, he moves into slot 1, and customers formerly in slot i move into slot $i+1$, $1 \leq i \leq n$. We also assume that if a customer arrives and finds N customers in the system, he eliminates the customer in slot k, $1 \leq k \leq N-1$, with probability α_k. If he decides not to eliminate one of these customers

https://doi.org/10.1017/S0269964813000041 Published online by Cambridge University Press

(which occurs with probability λ_{N-1}/λ_0), he joins slot 1, pushing the customer formerly in slot i to slot $i+1$, $1 \le i \le N$, so that the customer formerly in slot N is eliminated due to overcrowding.

Finally, the server always processes the work of the customer in the ith slot at a rate $\mu_i - \mu_{i-1}$ ($\mu_0 = 0$), so that the total rate at which the work is processed while there are n customers in the system is simply μ_n. When a customer in slot i completes its service, the customer formerly in slot $j > i$ moves to slot $j-1$, for all slots $j > i$ containing a customer at the service completion instant.

It is not difficult to see that Q is equal in distribution to the queue-length process of the knockout queue. Moreover, the knockout queue has many useful properties: (i) the arrival process of customers to the knockout queue is a homogeneous Poisson process with rate λ_0, (ii) the sojourn time of a customer arriving at time s depends only on the amount of work he brings to the system, along with the work brought by customers that arrive after time s, and (iii) the sojourn time of a customer is always equal in distribution to τ, which represents the amount of time it takes a CTMC $S := \{S(t); t \ge 0\}$ with generator $R = [r(i,j)]_{0 \le i,j \le N+1}$ to reach the absorbing set $\{0, N+1\}$, where R is of the form

$$r(i,j) = \begin{cases} \lambda_{i-1} - \lambda_i, & 1 \le i \le N-1, j = N+1; \\ \mu_{i-1} - \mu_0, & 1 \le i \le N, j = i-1; \\ \lambda_i, & 1 \le i \le N, j = i+1; \\ \mu_i - \mu_{i-1}, & 1 \le i \le N, j = 0; \\ 0, & \text{otherwise.} \end{cases}$$

To see why the generator of S is of this form, suppose that at a given time, a particular customer is in slot i. The customer will move to slot $i+1$ if someone arrives and either eliminates a customer in a higher-labeled slot, or chooses not to eliminate anyone, and this occurs with rate $\lambda_0(\lambda_i/\lambda_0) = \lambda_i$. Next, the customer will move to slot $i-1$ if a customer in one of the lower-labeled slots completes its service, and this occurs with rate $\mu_{i-1} - \mu_0$. Finally, from state i the customer could be eliminated from the system by an arrival, or it could complete its service: these events occur with rates $\lambda_0\alpha_i = \lambda_{i-1} - \lambda_i$ and $\mu_i - \mu_{i-1}$, respectively. Note that, S reaching state $N+1$ first corresponds to a customer being eliminated, while S reaching state 0 first corresponds to a service completion.

2.1. The Time-Dependent First Moments

We are now ready to state our first main result.

THEOREM 2.1: *The time-dependent first moments of the knockout queue are as follows: for each $t \ge 0$,*

$$E_k[Q(t)] = \sum_{j=1}^{k} P_j(\tau > t) + \lambda_0 E_1[\tau] P_1(R_\tau \le t). \tag{1}$$

Here P_k represents a conditional probability, given $Q(0) = k$ if our probability is in terms of Q, or given $S(0) = k$ if our probability is in terms of τ, and R_τ represents the equilibrium version, or residual, of τ. Recall that for a non-negative random variable X, its residual R_X is a random variable whose cumulative distribution function is of the form

$$P(R_X \le t) = \frac{1}{E[X]} \int_0^t P(X > s) ds,$$

for each $t \ge 0$.

https://doi.org/10.1017/S0269964813000041 Published online by Cambridge University Press

PROOF: Fix a $t \geq 0$. To derive the first moment, recall from the transient Little's law [8] that for each $0 \leq k \leq N$ integer,

$$E_k[Q(t)] = \sum_{j=1}^{k} P_j(\tau > t) + \int_0^t \mathcal{P}_s(W(s) > t - s)\lambda_0 ds,$$

where $\mathcal{P}_s$ represents the time-dependent Palm probability induced by the Poisson arrival process with rate λ_0, and $W(s)$ is the sojourn time of the customer arriving at time s: recall that under $\mathcal{P}_s$, an arrival occurs at time s with probability one (we refer readers who are interested in learning more about Palm probabilities to Kallenberg [10]).

Finally, the dynamics of the knockout model imply that for each $s > 0$, $\mathcal{P}_s(W(s) > t - s) = P_1(\tau > t - s)$, since the sojourn time of the customer arriving at time s is only dependent on the customers arriving after time s. ■

It is interesting to note that when the knockout model is empty at time zero, we have that

$$E_0[Q(t)] = \lambda_0 E_1[\tau] P_1(R_\tau \leq t)$$

which leads to the following corollary.

COROLLARY 2.1: *$E_0[Q(t)]$ is a concave function in t.*

PROOF: It suffices to note that the derivative of $E_0[Q(t)]$ is decreasing, proving concavity. ■

This concavity result can also be found on page 162 of Lindvall [11], where it is proven via coupling techniques.

2.2. The Time-Dependent Balking Probability

We next focus on deriving the probability $\mathcal{P}_{k,t}(B(t))$, where $B(t)$ is the event that the customer arriving at time t in the FCFS system decides not to enter the system, and $\mathcal{P}_{k,t}$ is the Palm probability induced by the Poisson arrival process with rate λ_0, given that k customers are present in the system at time zero: again, recall that under $\mathcal{P}_{k,t}$, a customer arrives at time t with probability one. Readers should note that the time-dependent balking probability given above is a natural quantity to study, despite it being expressed as a Palm probability. In fact, for some models it can be expressed in terms of the underlying measure P. For example, note that when $\lambda_n = \lambda_0$ for $1 \leq n \leq N - 1$, a customer only balks if he observes N customers already present upon arrival, and so by letting $Q(t-) = \lim_{s \uparrow t} Q(s)$, we have that

$$\mathcal{P}_{k,t}(B(t)) = \mathcal{P}_{k,t}(Q(t-) = N) = P_k(Q(t) = N),$$

where the last equality follows from the transient PASTA property; see [6] for details regarding transient PASTA. In the $M/M/N/N$ loss model setting, $P_k(Q(t) = N)$ is carefully studied in [5,12,15].

We state our expression for $\mathcal{P}_{k,t}(B(t))$ in terms of random variables τ_i, where τ_i represents the first time the S process hits state i.

https://doi.org/10.1017/S0269964813000041 Published online by Cambridge University Press

THEOREM 2.2: *For almost-every (with respect to Lebesgue measure) $t \geq 0$, and each $0 \leq k \leq N$, we have that*

$$\mathcal{P}_{k,t}(B(t)) = \frac{1}{\lambda_0}\frac{d}{dt}\left[\sum_{j=1}^{k} P_j(\tau_{N+1} \leq t, \tau_{N+1} < \tau_0)\right] + P_1(\tau_{N+1} \leq t, \tau_{N+1} < \tau_0). \quad \textbf{(2)}$$

PROOF: We first compute the expected number of knockouts occurring in the interval $(0,t]$, where we denote the number of knockouts occurring in $(0,t]$ as $K(0,t]$. Observe first that a customer found in slot j at time 0 is knocked-out at or before time t with probability $P_j(\tau_{N+1} \leq t, \tau_{N+1} < \tau_0)$. Moreover, for a customer arriving at time s, the probability he is knocked-out at or before time t is $P_1(\tau_{N+1} \leq t-s, \tau_{N+1} < \tau_0)$. Applying the Campbell–Mecke formula (see [8], for example), we see that the expected number of knockouts occurring in $(0,t]$ is

$$E[K(0,t]] = \sum_{j=1}^{k} P_j(\tau_{N+1} \leq t, \tau_{N+1} < \tau_0) + \lambda_0 \int_0^t P_1(\tau_{N+1} \leq t-s, \tau_{N+1} < \tau_0)ds.$$

Furthermore, the expected number of customers suffering a knockout in $(0,t]$ is also the expected number of customers that balk upon arrival in $(0,t]$ in the FCFS system, where the server operates at rate μ_n whenever there are n customers in the system, and where customer elimination is replaced by ordinary balking, with balking occurring with probability $1-\lambda_n/\lambda_0$ if there are n customers observed in the system upon arrival. Applying the Campbell–Mecke formula with this observation yields

$$E[K(0,t]] = \lambda_0 \int_0^t \mathcal{P}_{k,s}(B(s))ds,$$

so for each $t \geq 0$, we have that

$$\lambda_0 \int_0^t \mathcal{P}_{k,s}(B(s))ds = \sum_{j=1}^{k} P_j(\tau_{N+1} \leq t, \tau_{N+1} < \tau_0) + \lambda_0 \int_0^t P_1(\tau_{N+1} \leq t-s, \tau_{N+1} < \tau_0)ds,$$

and taking derivatives of both sides with respect to t completes the proof. ■

3. EXAMPLES

The examples given below illustrate that in some settings, our first moment expressions and balking probabilities simplify further. We assume that $Q(0) = 0$ in each of the examples given below, however readers should note that similar observations may be made for any possible initial condition.

3.1. The $M/M/1/1$ Queue

We first consider the $M/M/1/1$ queue, where customers arrive according to a Poisson process with rate λ, each customer brings with it an exponentially distributed amount of work with unit rate, and work is processed with rate μ. Thinking of this as a knockout model, we see that the sojourn time of a typical customer is the minimum of the amount of work it brings, and the time remaining after its arrival to the system until the next arrival.

https://doi.org/10.1017/S0269964813000041 Published online by Cambridge University Press

In other words, its sojourn time is exponential with rate $\lambda+\mu$. Assuming the system is empty at time zero, we see from (1) that

$$P(Q(t)=1 \mid Q(0)=0) = E_0[Q(t)] = \lambda \int_0^t e^{-(\lambda+\mu)(t-s)} ds = \frac{\lambda}{\lambda+\mu}(1-e^{-(\lambda+\mu)t}),$$

which agrees with the well-known time-dependent results for the $M/M/1/1$ system (see e.g., Ross [14]).

3.2. The $M/M/N/N$ Loss Model

Our next example focuses on the first moment of the $M/M/N/N$ loss model $\{Q(t); t \geq 0\}$, with arrival rate λ and service rate μ for each customer. Thinking of the loss model as a knockout queue, we see that an arbitrary arrival to the system leaves the system whenever either (i) he completes service or (ii) an arrival occurs while he, along with $N-1$ customers that arrived after him, are still in the system. Hence, in this case (1) yields

$$\begin{aligned} E_0[Q(t)] &= \lambda \int_0^t P_0(\tau_{M/M/\infty,N} > t-s, e_\mu > t-s) ds \\ &= \lambda \int_0^t P_0(\tau_{M/M/\infty,N} > t-s) e^{-\mu(t-s)} ds, \end{aligned}$$

where $\tau_{M/M/\infty,N}$ is the amount of time it takes a $M/M/\infty$ queue with arrival rate λ and service rate μ to reach state N, and e_μ is an exponential random variable with rate μ.

3.3. The $M/M/1/N$ Queue

Additional computations of the time-dependent loss probabilities can be performed for the $M/M/1/N$ queue, thanks to martingale theory. Suppose $\{A(t); t \geq 0\}$ and $\{D(t); t \geq 0\}$ are two independent homogeneous Poisson processes with rates λ and μ, respectively, and set $X(t) = A(t) - D(t)$. Furthermore, set

$$Z(t) = e^{qX(t)-f(q)t},$$

where for each $q \geq 0$,

$$f(q) = -\lambda(1-e^q) - \mu(1-e^{-q}).$$

Choosing f in this manner makes $\{Z(t); t \geq 0\}$ a martingale with respect to itself.
Fix two integers a, b, where $b < 0 < a$, and define $T_{a,b} = \inf\{t \geq 0 : X(t) \in \{a, b\}\}$, $T_a = \inf\{t \geq 0 : X(t) = a\}$ and similarly for T_b. An application of the optional sampling theorem yields

$$\begin{aligned} 1 &= E[e^{q(X(T_{a,b}))-f(q)T_{a,b}}] \\ &= E[e^{q(X(T_{a,b}))-f(q)T_{a,b}} \mathbf{1}(T_a < T_b)] + E[e^{q(X(T_{a,b}))-f(q)T_{a,b}} \mathbf{1}(T_a > T_b)] \\ &= e^{qa} E[e^{-f(q)T_a} \mathbf{1}(T_a < T_b)] + e^{qb} E[e^{-f(q)T_b} \mathbf{1}(T_b < T_a)] \\ &= e^{qa} E[e^{-f(q)T_a} \mathbf{1}(T_a < T_b)] + e^{qb} \left[E[e^{-f(q)T_{a,b}}] - E[e^{-f(q)T_a} \mathbf{1}(T_a < T_b)] \right], \end{aligned}$$

and after a bit of algebra, we see that

$$E[e^{-f(q)T_a} \mathbf{1}(T_a < T_b)] = \frac{e^{qb} E[e^{-f(q)T_{a,b}}] - 1}{e^{qb} - e^{qa}}.$$

https://doi.org/10.1017/S0269964813000041 Published online by Cambridge University Press

This expression can be written in a more convenient form: recall that the inverse function $\Phi : [0, \infty) \to \mathbb{R}$ of f exists, in that for each $q \geq 0$, $\Phi(f(q)) = f(\Phi(q)) = q$. In fact, Φ can be expressed in closed-form: for each $q \geq 0$,

$$\Phi(q) = \ln \left[\frac{(q + \lambda + \mu) + \sqrt{(q + \lambda + \mu)^2 - 4\lambda\mu}}{2\lambda} \right].$$

Hence,

$$E[e^{-qT_a} \mathbf{1}(T_a < T_b)] = \frac{e^{\Phi(q)b} E[e^{-qT_{a,b}}] - 1}{e^{\Phi(q)b} - e^{\Phi(q)a}}.$$

We are now ready to compute $P_0(Q(e_q) = N)$, where e_q is an exponential random variable with rate $q > 0$, independent of Q. We focus here on the case where the system is empty at time zero, but the same approach may be used for all other possible initial states. By making use of (2), we see that

$$\begin{aligned}
P_0(Q(e_q) = N) &= \int_0^\infty P_0(Q(t) = N) q e^{-qt} dt \\
&= \int_0^\infty P_1(\tau_{N+1} \leq t, \tau_{N+1} < \tau_0) q e^{-qt} dt \\
&= P_1(\tau_{N+1} \leq e_q, \tau_{N+1} < \tau_0) \\
&= E_1[e^{-q\tau_{N+1}} \mathbf{1}(\tau_{N+1} < \tau_0)] \\
&= E_0[e^{-qT_N} \mathbf{1}(T_N < T_{-1})] \\
&= \frac{e^{-\Phi(q)} E[e^{-qT_{-1,N}}] - 1}{e^{-\Phi(q)} - e^{\Phi(q)N}} \\
&= \frac{E[e^{-qT_{-1,N}}] - e^{\Phi(q)}}{1 - e^{\Phi(q)(N+1)}}.
\end{aligned}$$

Note that with this approach, we do not need to compute any other terms of the form $P_i(Q(e_q) = j)$: this would be the case if we were to compute these probabilities via the Kolmogorov Forward equations.

3.4. The $M/M/N/N$ Queue

We now further study the time-dependent loss probabilities of the $M/M/N/N$ loss model. If we again make use of the fact that the probability a customer is knocked out of the system is equal to the probability that an external loss model with N servers reaches level N before the customer completes service, starting from state zero, we find that

$$\begin{aligned}
P_0(Q(t) = N) &= P_1(\tau_{N+1} \leq t, \tau_{N+1} \leq \tau_0), \\
&= P_0 \left(\sup_{0 \leq s \leq \min(t, e_\mu)} Q_\infty(s) \geq N \right),
\end{aligned}$$

where Q_∞ represents the queue-length process of a $M/M/\infty$ queue with arrival rate λ and service rate μ that is independent of e_μ, with e_μ being thought of as the amount of time it

https://doi.org/10.1017/S0269964813000041 Published online by Cambridge University Press

takes to process the customer's amount of work. Integrating with respect to an exponential density with rate $q > 0$ yields, after an application of (2),

$$\begin{aligned} P_0(Q(e_q) = N) &= P_1(\tau_{N+1} \le e_q, \tau_{N+1} \le \tau_0) \\ &= P_0\left(\sup_{0\le s\le \min(e_q,e_\mu)} Q_\infty(s) \ge N\right) \\ &= E_0[e^{-(q+\mu)\tau_{M/M/\infty,N}}] \\ &= \frac{\sum_{m=0}^{N}\binom{N}{m}\frac{(\rho)^N(-1)^m}{(N)!}\frac{q+\mu}{q+(m+1)\mu}M\left(1,\frac{q}{\mu}+m+2,-\rho\right)}{\sum_{j=0}^{N}\sum_{m=0}^{2(N-j)}\binom{N}{j}\binom{2(N-j)}{m}\frac{(\rho)^{N-j}(-1)^m}{(N-j)!}\frac{q+\mu}{q+(j+m+1)\mu}} \\ &\qquad \times M\left(1,\frac{q}{\mu}+j+m+2,-\rho\right), \end{aligned}$$

where the last equality follows from Lemma 3.3 of [9], with $\rho = \lambda/\mu$. Here, M represents the Kummer function, where

$$M(a,b,z) = \sum_{n=0}^{\infty}\frac{(a)_n z^n}{(b)_n n!},$$

with $(a)_0 = 1$, and for $n \ge 1$, $(a)_n = (a)(a+1)\cdots(a+n-1)$. Again, we note as before that no other probabilities of the form $P_i(Q(e_q) = j)$ need to be computed in order to compute the loss probability.

We close by noting that for this particular queueing system, an alternative method can be used to derive expressions for the time-dependent loss probability. It is well-known that by the Campbell-Mecke formula (see also [5])

$$E_0[Q(t)] = \lambda\int_0^t e^{-\mu(t-s)}P_0(Q(s-) \le N-1)ds,$$

implying

$$\lambda\int_0^t e^{\mu s}P_0(Q(s-) \le N-1)ds = \lambda e^{\mu t}\int_0^t P_1(\tau > s)ds.$$

Taking derivatives of both sides with respect to t gives

$$\lambda e^{\mu t}P_0(Q(t) \le N-1) = \lambda e^{\mu t}P_1(\tau > t) + \lambda\mu e^{\mu t}\int_0^t P_1(\tau > s)ds$$

or

$$P_0(Q(t) \le N-1) = P_1(\tau > t) + \mu\int_0^t P_1(\tau > s)ds = P_1(\tau > t) + \mu E_1[\tau]P_1(R_\tau \le t),$$

where R_τ is itself also a phase-type random variable (see Chapter 5 of Neuts [13]). Rewriting this expression gives the following alternative expression for the loss probability:

$$P_0(Q(t) = N) = P_1(\tau \le t) - \mu E_1[\tau]P_1(R_\tau \le t).$$

https://doi.org/10.1017/S0269964813000041 Published online by Cambridge University Press

References

1. Abate, J. & Whitt, W. (1987). Transient behavior of regulated Brownian motion, I: starting at the origin. *Advances in Applied Probability* **19**, 560–598.
2. Abate, J. & Whitt, W. (1987). Transient behavior of regulated Brownian motion, II: non-zero initial conditions. *Advances in Applied Probability* **19**, 599–631.
3. Abate, J. & Whitt, W. (1987). Transient behavior of the $M/M/1$ queue: starting at the origin. *Queueing Systems* **2**, 41–65.
4. Abate, J. & Whitt, W. (1988). Transient behavior of the $M/M/1$ queue via Laplace transforms. *Advances in Applied Probability* **20**, 145–178.
5. Abate, J. & Whitt, W. (1995). Calculating transient characteristics of the Erlang loss model via numerical transform inversion. *Stochastic Models* **14**, 663–680.
6. Fralix, B.H., Riaño, G., & Serfozo, R.F. (2007). Time-dependent Palm probabilities and queueing applications. Unpublished technical report. Available from http://www.eurandom.nl/reports/2007/041-report.pdf.
7. Fralix, B.H. (2012). On the time-dependent moments of Markovian queues with reneging. *Queueing Systems*, to appear.
8. Fralix, B.H. & Riaño, G. (2010). A new look at transient versions of Little's law, and $M/G/1$ preemptive-last-come-first-served queues. *Journal of Applied Probability* **47**, 459-473.
9. Fralix, B.H., van Leeuwaarden, J.S.H., & Boxma, O.J. (2012). Factorization Identites for a general class of reflected processes. Available from http://www.eurandom.nl/reports/2011/024- report.pdf.
10. Kallenberg, O. (1983). *Random Measures*. Berlin: Akademie-Verlag.
11. Lindvall, T. (2002). *Lectures on the Coupling Method*. Mineola, New York: Dover Publications, Inc.
12. Mandjes, M. & Ridder A. (2001). A large deviations approach to the transient of the Erlang loss model. *Performance Evaluation* **43**, 181–198.
13. Neuts, M.F. (1989). *Structured stochastic matrices of $M/G/1$-type and their applications*. New York: Marcel Dekker, Inc.
14. Ross, S.M. (1990). *Stochastic Processes*, 2nd ed. New York: Wiley.
15. Virtamo, J. & Aalto, S. (1998). Calculation of time-dependent blocking probabilities. Proceedings of the ITC Sponsored St. Petersburg Regional International Teletraffic Seminar Teletraffic Theory as a Base for QoS: Monitoring, Evaluation, Decisions 365–375. Available from: http://www.netlab.tkk.fi/tutkimus/cost257/.

https://doi.org/10.1017/S0269964813000041 Published online by Cambridge University Press